



## Online Tools and Training for Access and Analysis of Restricted Government Data Files



IASSIST 2016  
Bergen, Norway  
3 June 2016



Cornell University

# CISER

Cornell Institute for Social and Economic Research  
A LEADER IN SOCIAL SCIENCE DATA AND COMPUTING

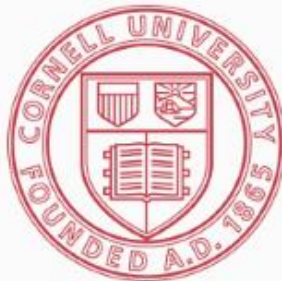


Cornell University

## CED<sup>2</sup>AR

Official Server - The Comprehensive Extensible Data Documentation and Access Repository

Supported By



Built With



## Target Audience

Researchers working with official statistical files

- With approved access to restricted files
- Applying for access to restricted files
- Public use files

## Facilitating Work With Official Restricted Access Statistical Files

- Discovery
- Metadata
- Practice Data Sets
- Training

## Discovery

### CED<sup>2</sup>AR Metadata Repository



DDI



## Facilitating Work With Official Restricted Access Statistical Files

- Discovery
- ***Metadata***
- Practice Data Sets
- Training

## Filter Codebooks



ACS PUMS ☐

CNSS ☐

ECF ☐

IPUMS-USA ☐

LBD ☐



NBER CES ☐

NHIS ☐

National QWI ☐

SSB ☐

SynLBD ☐

## Search

Searching all codebooks. No filters active.



Rectangular Snip

Search

[Advanced Search](#)

Show  variables



# Compare Variables Across Files

CED<sup>2</sup>AR

Development Server - The Comprehensive Extensible Data Documentation and Access Repository

[Search Variables](#) [Browse Variables ▾](#) [Browse by Codebook](#) [Documentation](#) [About](#)

## Comparison Overview



Name	Codebook	Label	Description
<a href="#">foreign_born</a>	<a href="#">SIPP Synthetic Beta v5.1</a>	Foreign Born	Immigrant Status, born in country other than U.S. Taken from wave 2 topical module (TM8730, TM8734, TM8709 1990-1993 panels; eprstate, ebrstate and rcitiznt 1996 panel; eprstate, ebrstate and tcitiznt 2001 panel; eprstate, ebrstate, citiz, and eborn ... <a href="#">more</a>
<a href="#">NATIVITY</a>	<a href="#">IPUMSUSA</a>	Foreign birthplace or parentage	NATIVITY indicates whether respondents were native-born or foreign-born; for native-born respondents, it indicates whether their mothers and/or fathers were native-born or foreign-born. NATIVITY is constructed from the IPUMS variables BPL, MBPL, and ... <a href="#">more</a>
<a href="#">MIGSP</a>	<a href="#">ACS 2009 Public Use File</a>	Migration recode - State or foreign country code	



## Compare Variable Values

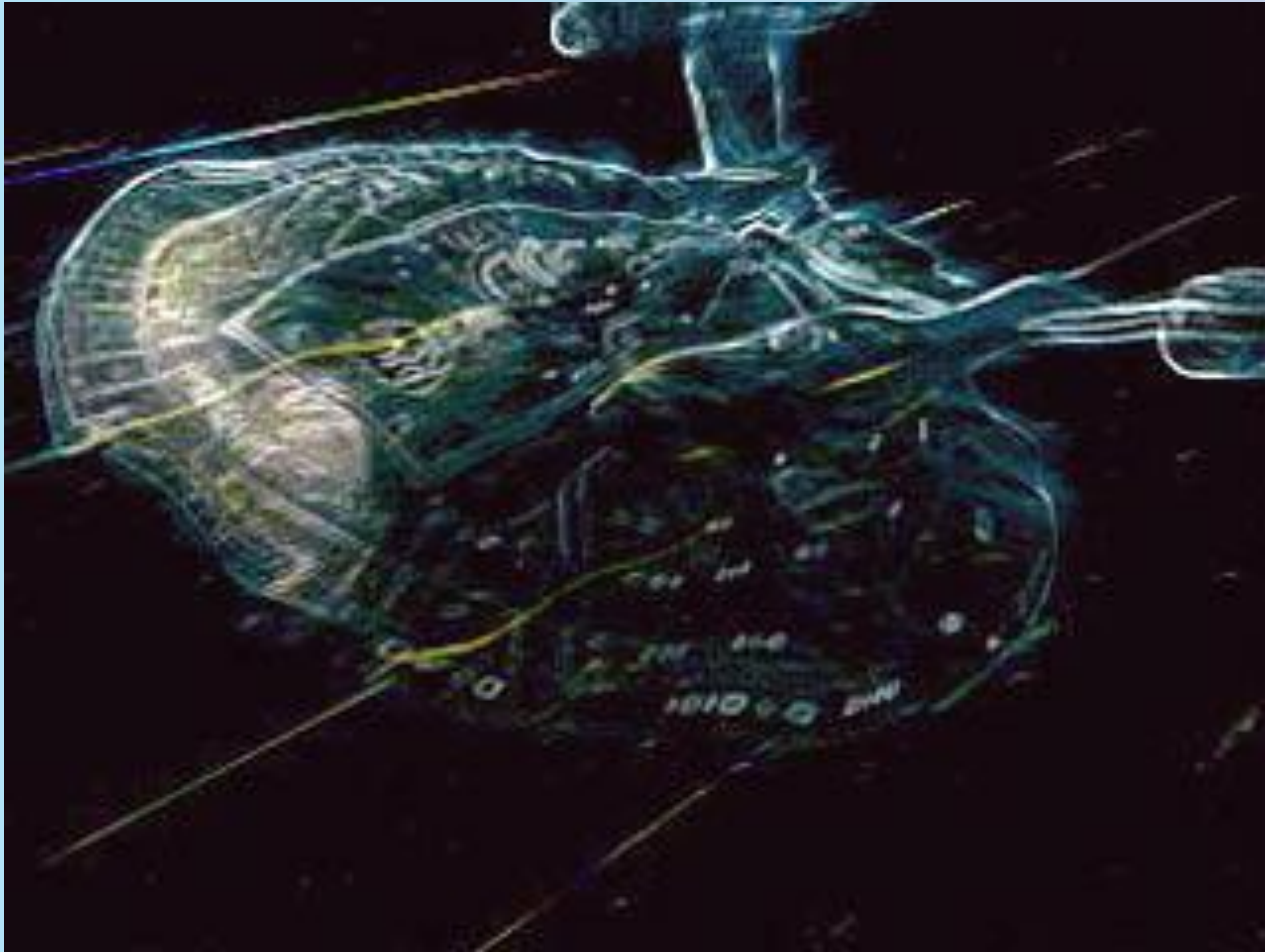
### Values

Values	ACS 2009 Public Use File (100 total)	MIGSP SIPP Synthetic Beta v5.1 (2 total)	foreign_born IPUMSUSA (6 total)	NATIVITY
Africa	✓	✗	✗	
Alabama/AL	✓	✗	✗	
Alaska/AK	✓	✗	✗	
Arizona/AZ	✓	✗	✗	
Arkansas/AR	✓	✗	✗	
Australia	✓	✗	✗	
Australian and New Zealand subregions, not specified, Oceani	✓	✗	✗	
Born in U.S.	✗	✓	✗	
Born in country other than U.S.	✗	✓	✗	
Brazil	✓	✗	✗	
California/CA	✓	✗	✗	
Canada	✓	✗	✗	
Caribbean and North America, not specified	✓	✗	✗	
Central America, not specified	✓	✗	✗	
China, Hong Kong and Paracel Islands	✓	✗	✗	
Colombia	✓	✗	✗	
Colorado/CO	✓	✗	✗	
Connecticut/CT	✓	✗	✗	
Cuba	✓	✗	✗	

# Values

Values	MIGSP	foreign_born	NATIVITY
	ACS 2009 Public Use File (100 total)	SIPP Synthetic Beta v5.1 (2 total)	IPUMSUSA (6 total)
Africa	✓	✗	✗
Alabama/AL	✓	✗	✗
Alaska/AK	✓	✗	✗
Arizona/AZ	✓	✗	✗
Arkansas/AR	✓	✗	✗
Australia	✓	✗	✗
Australian and New Zealand subregions, not specified, Oceani	✓	✗	✗
Born in U.S.	✗	✓	✗
Born in country other than U.S.	✗	✓	✗
Brazil	✓	✗	✗
California/CA	✓	✗	✗
Canada	✓	✗	✗
Caribbean and North America, not specified	✓	✗	✗
Central America, not specified	✓	✗	✗

## Confidentiality and Cloaking in Metadata



- Marking elements with different restrictions

Select what sub-elements to mark

☐ Select All

☐ Mean

☐ Median

☐ Mode

☐ Valid

☐ Invalid

☒ Min

☒ Max

☐ Standard Deviation

☐ Other Summary Statistics

☐ Values

☒ Value Frequencies

☒ Value Percentages

☒ Value Crosstabs

☒ Other Value Statistics

☐ Label

☐ Notes

Select what access level to apply, then check which variables to apply to. Finally, click changes levels.

restricted ▼

Change Levels

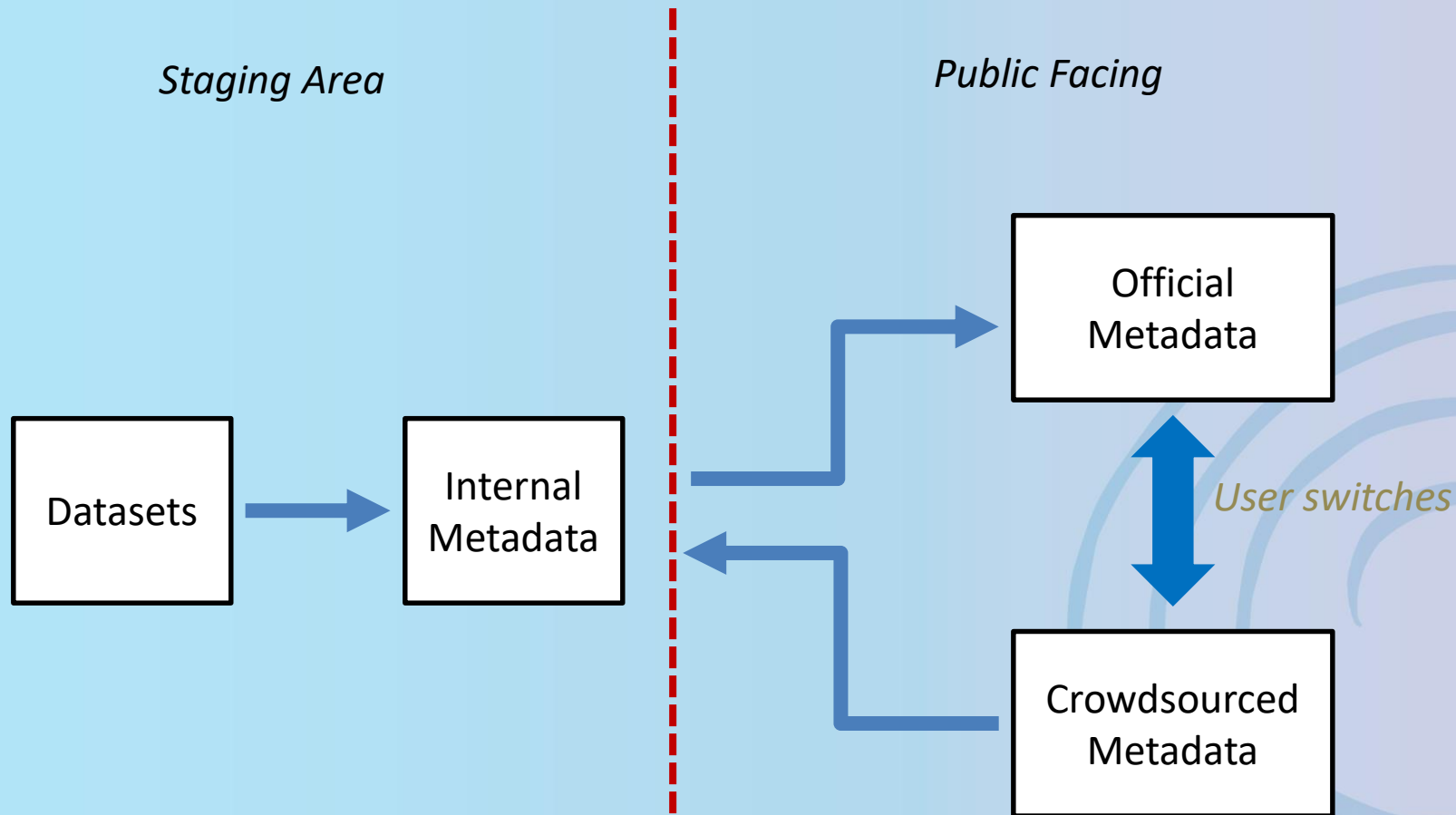
	Variable Name	Label	Top Access Level
<input checked="" type="checkbox"/>	afdc_MN	Indicator for receipt of AFDC or TANF benefits	released
<input checked="" type="checkbox"/>	afdcamt_MN	Amount of AFDC received	released
<input type="checkbox"/>	birthdate	Date of Birth	released
<input type="checkbox"/>	current_enroll_coll	Currently Enrolled in College	released
<input type="checkbox"/>	current_enroll_hs	Currently Enrolled in HS (or less)	released



## Crowdsourcing and Curated Metadata



# CISER Basic Information Flow



## Combining Knowledge: Merging

- Curators are given an interface to merge crowdsourced documentation with official

### Merge Variables

The following variables have changed:

cur\_endmar

birthdate

Continue

## Combining Knowledge: Merging

current\_enroll\_coll

### Crowdsourced Documentation

Variable Name	current_enroll_coll
Label	Currently Enrolled <u>in College</u>
Codebook	SIPP Synthetic Beta v6
Concept	
Concept Vocabulary	
Concept Vocabulary URI	
Type	numeric
Files	

### Official Documentation

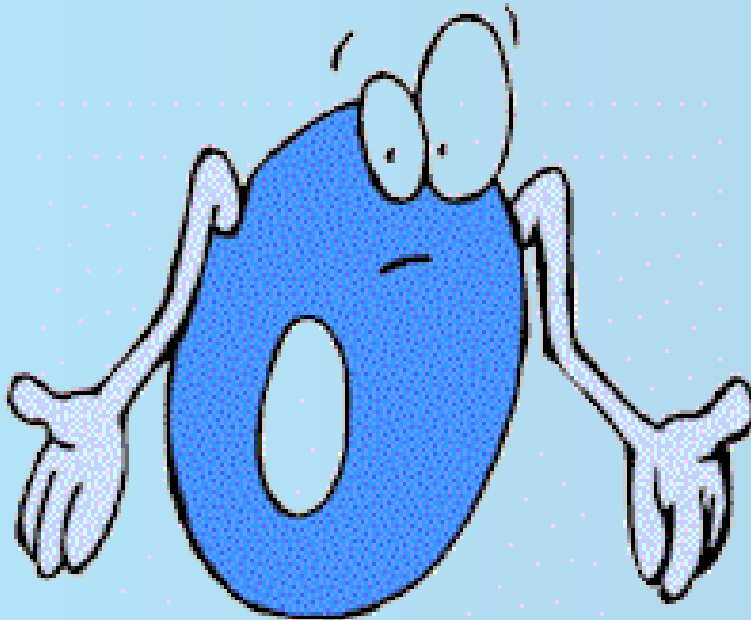
Variable Name	current_enroll_coll
Label	<input type="checkbox"/> Use crowdsourced <input type="checkbox"/> Use original         Currently Enrolled
Codebook	SIPP Synthetic Beta v6
Concept	
Concept Vocabulary	
Concept Vocabulary URI	
Type	numeric
Files	



## Facilitating Work With Official Restricted Access Statistical Files

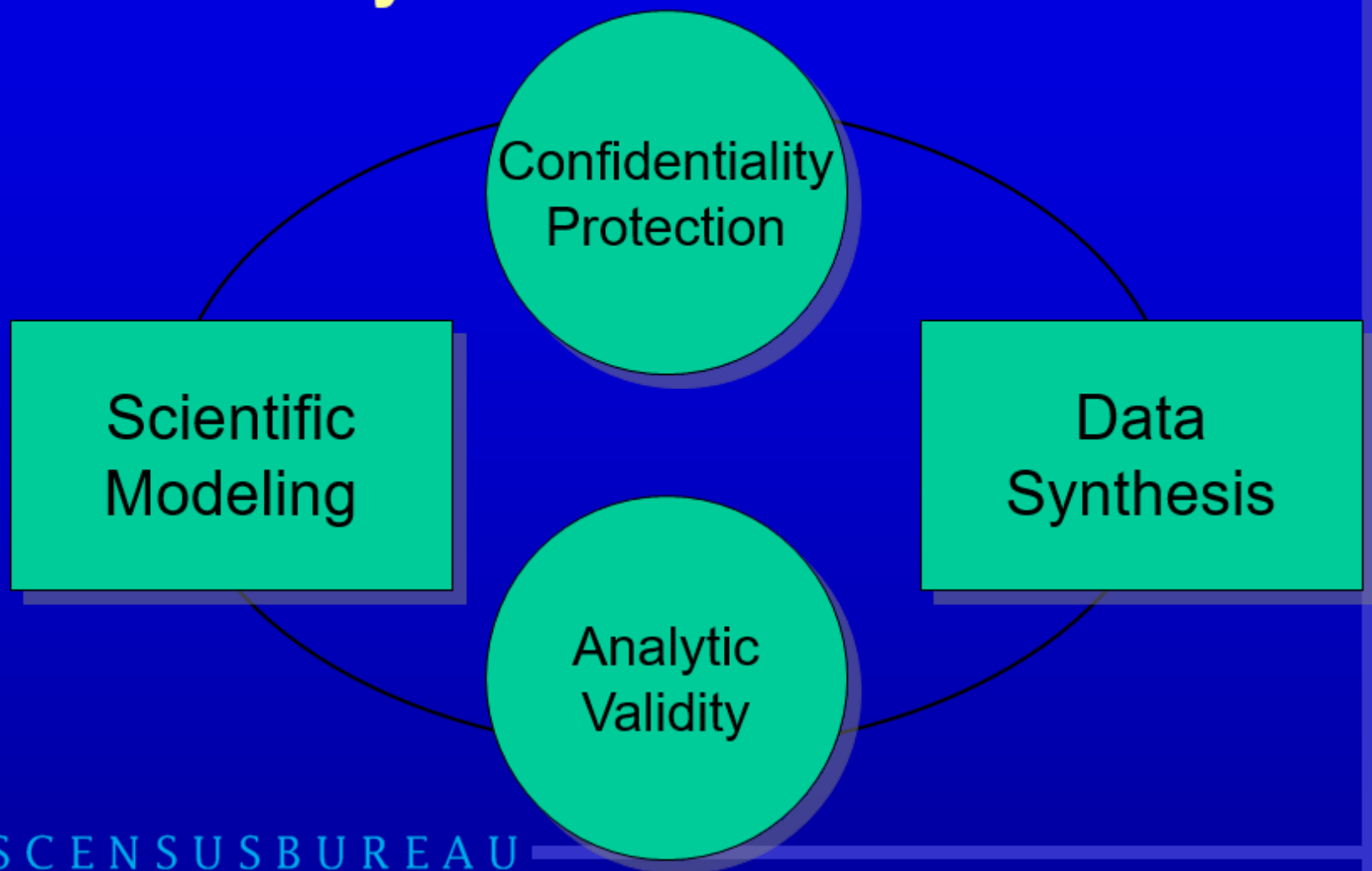
- Discovery
- Metadata
- ***Practice Data Sets***
- Training

## Zero Observation Data Sets



```
data likeshoes;  
set sasuser.shoes  
(obs=0);  
run;
```

# The Research – Synthetic Data Feedback Cycle





Cornell University

SEARCH CORNELL:

go



Pages



People

more options



## VirtualRDC @ Cornell

Synthetic Data Server, Econ Compute Cluster, and more

INFORMATION ✕ ECCO ✕ SYNTHETIC DATA SERVER ✕ DATA ✕ DOCUMENTATION ✕ HELP ✕ LABOR DYNAMICS INSTITUTE

### Access to ECCO

Economics Compute Cluster (ECCO):

- Account request procedures
- Usage instructions
- Short tutorial

### Synthetic Data Server

Info on Synthetic Data Server (SDS):

- Account request procedures
- Codebooks on CED<sup>2</sup>AR
- Short tutorial

### Quicklinks

- Data download
- Globus Online
- Login to compute nodes

## Synthetic Data Server

### On this page

- Available data
- In this section
- History
- Funding acknowledgement
- Bibliography

The Synthetic Data Server (SDS) was set up to provide early access to new synthetic data products by the U.S. Census Bureau. These datasets are made available to interested researchers in a controlled environment, prior to a more generalized release.

## Available data

At present, two datasets are made available on this server:

- **SIPP Synthetic Beta (SSB)**
- **Synthetic LBD (SynLBD)**

How to request access to the data is described on the **next page**. Access requests are reviewed for feasibility, but not otherwise restricted.

### Get us on your mobile device



### Site Navigation

- Documentation
- Help (ECCO, SDS, VirtualRDC)
- Workshops and Sessions
- Economics Compute Cluster Organization (ECCO)
- Synthetic Data Server
  - Step 1 - Requesting access to SDS

## CED<sup>2</sup>AR

Official Server - The Comprehensive Extensible Data Documentation and Access Repository

[Search Variables](#)

[Browse Variables](#) ▾

[Browse by Codebook](#)

[Documentation](#)

[About](#)



You are viewing the official metadata. View crowdsourced [contributions](#).

[CED2AR](#) / SIPP Synthetic Beta v6.02

## SIPP Synthetic Beta v6.02



[SAS](#)

[Stata](#)




[View Variables](#) (123 variables)


Last update to metadata: 2015-11-24 10:05:15 (upload date)

Document Date: November 12, 2015

Codebook prepared by: Cornell NSF-Census Research Network

## ▼ Datasets

ssb\_v6\_0\_2\_syntheticK\_M.sas7bdat <http://www.census.gov/programs-surveys/sipp/methodology/sipp-synthetic-beta-data-product.html>  ( SAS )

ssb\_v6\_0\_2\_syntheticK\_M.dta <http://www.census.gov/programs-surveys/sipp/methodology/sipp-synthetic-beta-data-product.html>  ( Stata )

Rectangular Snip

## ▼ Terms of Use

- Access Levels
- Access Restrictions (Default)
- Access Requirements
- Access Conditions
- Access Permission Requirements
- Citation Requirements
- Disclaimer

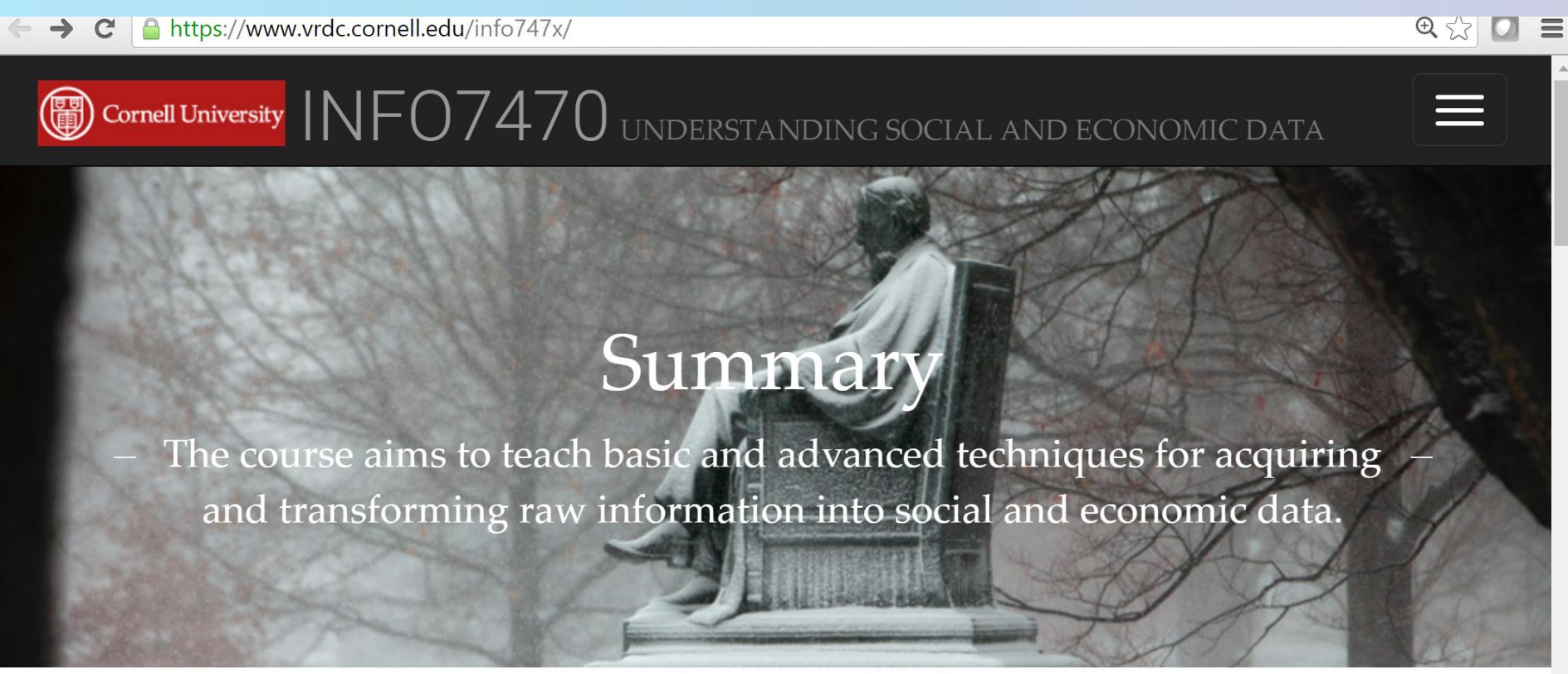
## Facilitating Work With Official Restricted Access Statistical Files

- Discovery
- Metadata
- Practice Data Sets
- ***Training***




## INFO7470

[www.vrdc.cornell.edu/info747x/](http://www.vrdc.cornell.edu/info747x/)

A screenshot of a web browser displaying the INFO7470 course page. The browser's address bar shows the URL 'https://www.vrdc.cornell.edu/info747x/'. The website header features the Cornell University logo on the left, the text 'INFO7470 UNDERSTANDING SOCIAL AND ECONOMIC DATA' in the center, and a hamburger menu icon on the right. The main content area has a background image of a statue of a man sitting in a chair, with the word 'Summary' overlaid in large white text. Below the title, there is a list item starting with a dash.

← → ↻ 🔒 <https://www.vrdc.cornell.edu/info747x/> 🔍 ☆ 📱 ☰

 Cornell University INFO7470 UNDERSTANDING SOCIAL AND ECONOMIC DATA ☰

# Summary

- The course aims to teach basic and advanced techniques for acquiring – and transforming raw information into social and economic data.



[Home](#)[News](#)[Events](#)[Documents](#)[Nodes](#)[Software](#)[Education](#)[HOME](#) > [SHORT-COURSE - ONE-DAY SIPP TRAINING WORKSHOP](#)[Window Split](#)

## Short-Course - One-day SIPP Training Workshop

The Michigan NCRN will sponsor a [one-day SIPP training workshop](#) at Census Bureau Headquarters in Suitland, Maryland on Friday, May 15. The workshop will focus on SIPP classic, with an introduction to the re-designed SIPP, and is open to analysts at government agencies and nonprofits in the Washington DC area. Please contact [MCRDC@umich.edu](mailto:MCRDC@umich.edu) for more information.

**Academic year:**

May-2015

**Primary and cross-listed field of class or course:**

[Economics](#)

[Other](#)

[Statistics](#)

**EXTERNAL LINK:**

[Training Announcement](#)  
[Calendar](#)  
[Announcement](#)

**NODES:**

[University of Michigan](#)

**AUTHOR(S):**

[Matthew Shapiro](#)  
[Margaret Levenstein](#)

**TYPE:**

[Short Course](#)

## Contributing to Productive and Successful Research





- NSF-Census Research Network: [ncrn.info/](http://ncrn.info/)
- Cornell Node: [ncrn.info/node/cornell-university](http://ncrn.info/node/cornell-university)
- CED<sup>2</sup>AR: [www2.ncrn.cornell.edu/ced2ar-web](http://www2.ncrn.cornell.edu/ced2ar-web)
- Synthetic Server: [www2.vrdc.cornell.edu/news/synthetic-data-server/](http://www2.vrdc.cornell.edu/news/synthetic-data-server/)
- INFO7470: [www.vrdc.cornell.edu/info747x/](http://www.vrdc.cornell.edu/info747x/)

Warren Brown [wab4@cornell.edu](mailto:wab4@cornell.edu)

